REMARKS

Applicants have amended claims 1, 9, and 17 and have cancelled and claims 4, 12, and 20. In view of these amendments and the following remarks, Applicants hereby request further examination and reconsideration of the application, and allowance of claims 1-3, 5-11, 13-19, and 21-24.

The Office has rejected claims 1, 2, 6, 8, 9, 10, 14 and 15 under 35 U.S.C. 102(b) as being anticipated by U.S. Patent No. 581,666 to Alshawi. Applicants respectfully request clarification from the Office regarding this citation and for purposes of this response assume the Office was referring to U.S. Patent No. 5,815,196 to Alshawi (Alshawi). The Office has also rejected claims 3 and 11 under 35 U.S.C. 103(a) as being unpatentable over in view of U.S. Patent No. 6,647,535 to Bozdagi et al. (Bozdagi), claims 4, 5, 7, 12, 13, 17, 18, 20, 21, 22, 23 and 24 are rejected under 35 U.S.C. 103(a) as being unpatentable over in view of International Application Number: PCT/US99/03028 to Kazeroonian et al.(Kazeroonian), claim 16 under 35 U.S.C. 103(a) as being unpatentable over Alshawi in view of U.S. Patent No. 5,900,908 to Kirkland et al. (Kirkland), and claim 19 under 35 U.S.C. 103(a) as being unpatentable over Alshawi in view of Kazeroonian and further in view of Bozdagi.

The Office asserts that Alshawi discloses in Fig 1., a video-based communications device (5,8) which provides segmentation of an AV signal (16) and the further processing of the audio [speech] portion of the signal to provide continuous speech-to-subtitles [speech-to-text] translation (19,21,22) that has the ability to overlay and display text subtitles onto AV signal in real-time [captioning](26). The Office asserts that Alshawi does not show the synchronizing of the text [caption] data with one or more cues in the AV signal. However, the Office asserts that Kazeroonian discloses for each scene, the textual information related to a particular scene can be determined using a speech recognizer [speech-to-text processor] on the audio portion of the signal [and which is executed on a computer with a stored recordable medium (Page 13, Line 33). The Office asserts that in highly dynamic real time video this indexing feature is important in synchronizing the AV signal to the shown textual [caption] data. The Office asserts that it would have been obvious to one of ordinary skill in the art at the time of invention to further modify Alshawi with the addition of cues/indexing to the AV signal to help synchronize text [caption] and AV signal data as taught by Kazeroonian in order to improve on the real time captioning system for AV signals.

Additionally, the Office asserts that Alshawi does not show a method of converting the audio portion of the signal to text data that checks whether the amount of caption data is greater than a threshold amount or an expiration time before the process of association occurs. However, the Office asserts that Bozdagi discloses the use of processing a multimedia document which summarizes the original video by placing representative static images and text into a web document for viewing (Col 2, 5). Additionally, the Office asserts that Bozdagi discloses that the device can control the number of representative images transferred to be displayed by the use of a threshold (Col. 5, Lines 45-55) and that time is used to check the change in intensity between representative images (Col. 6, 7). The Office asserts it would have been obvious to one of ordinary skill at the time of the invention to modify Alshawi by the use of parameters such as caption amount and time threshold as taught by Bozagi et al. that show the benefits of the association of text and images for multimedia documents which may include AV signals.

Alshawi, Bozdagi, Kazeroonian, and Kirkland, alone or in combination, do not disclose or suggest, "wherein the associating further comprises synchronizing the caption data with one or more cues in the AV signal" as recited in claims 1 and 17 or "wherein the signal combination processing system synchronizes the caption data with one or more cues in the AV signal" as recited in claim 9. As the Office has acknowledged, Alshawi does not show the synchronizing of the text [caption] data with one or more cues in the AV signal. However, contrary to the Office's assertions Alshawi in view of Kazeroonian does not teach or suggest the claimed invention. Similarly, neither Bozdagi nor Kirkland, alone or in combination with the other cited references, teach or suggest the claimed invention.

Kazeroonian discloses a method and system for searching for and generally matching text from outside sources with an audio-video clip, but this text is not originally derived from the audio signal in the audio-video clip and thus has no direct timing correlation with the original audio-video clip. More specifically, as described at page 9, lines 25-35 in Kazeroonian, "An example of this type of correlation of sources is an audio-video clip of a weather forecast being matched to a textual weather forecast. This match could later be used to present to a user at a client computer 109 the textual weather forecast side-by-side with the audio-video weather forecast broadcast. Based on this type of correlation, news stories that

fall in the same category as a video clip can also be shown side by side. The user is therefore able to see retrieved video together with related news stories." The Office has cited to page 13, lines 29-32 in Kazeroonian, but all this section states is that the processor 240 matches the closed-caption text to available text content sources from other outside sources, such as matching the textual weather forecast with the audio-video weather forecast discussed above. Again, the text obtained from outside sources is not originally derived from the audio content or any other portion of the audio-video clip and thus has no direct timing correlation with the audio-video clip. There is simply no discussion in Kazeroonian or any of the other cited references on how to synchronize data which is converted from an audio signal in the audio-video clip back with the video signal associated with the converted audio signal. Thus, if as proposed by the Office, Alshawi is considered in view of Kazeroonian, at most the cited references would only disclose generally associating text from separate outside sources with the videophone conversation.

With the present invention, the caption data is synchronized with the cues in the audio-video signal so that the words coming out of the speaker's mouth on the audio-video signal correspond with the caption data being displayed. As a result, a hearing impaired individual can read the caption data and/or lip read from the speaker's mouth because the caption data and the audio signal have been synchronized. This synchronization of the caption data with the audio-video signal substantially enhances the ability of a hearing impaired individual to comprehend and retain the content of the audio-video signal. One example of the types of cues which can be added to the video signal to accomplish this is described in paragraph 20 in the above-identified patent application.

Accordingly, in view of the foregoing amendments and remarks, the Office is respectfully requested to reconsider and withdraw the rejection of claims 1, 9, and 17. Since claims 2-3 and 5-8 depend from and contain the limitations of claim 1, claims 10-11 and 13-16 depend from and contain the limitations of claim 9, and claims 18-19 and 21-24 depend from and contain the limitations of claim they are patentable in the same manner as claims 1, 9, and 17.

Alshawi, Bozdagi, Kazeroonian, and Kirkland, alone or in combination, also do not disclose or suggest, "determining a first amount of data in the caption data . . .

providing the caption data for the associating when the first amount is greater than a threshold amount or when a first period of time has expired" as recited in claims 3 and 19 or "wherein the speech-to-text processing system further comprises a counter that determines a first amount of data in the caption data and a timer that determines when a first period of time has expired, wherein the speech-to-text processing system providing the caption data for the associating when the first amount is greater than a threshold amount or when timer indicates that the first period of time has expired" as recited in claim 11. As the Office has acknowledged, Alshawi does not show a method of converting the audio portion of the signal to text data that checks whether the amount of caption data is greater than a threshold amount or an expiration time before the process of association occurs. However, contrary to the Office's assertions Alshawi in view of Bozdagi does not teach or suggest the claimed invention. Similarly, neither Kazeroonian nor Kirkland, alone or in combination with the other cited references, teach or suggest the claimed invention.

Bozdagi relates to a method and system for automatically parsing a video data signal to identify a subset of representative frames from a set of frames. More specifically, as disclosed at col. 5, lines 11-25 in Bozdagi, an image significance determiner 40 decides whether a selected frame within a segment should be kept as a representative image for that segment. Bozdagi discloses at col. 4, lines 44-67 the use of a frame difference determiner that computes the difference between two consecutive frames on a pixel by pixel basis which is then used to select a representative frame. Additionally, Bozdagi discloses at col. 6, lines 13-35, identifying a frame as a representative frame if the change in intensity between consecutive frames is determined to be greater than a threshold. Further, Bozdagi discloses at col. 7, lines 33-35, an extended time lapse between command data can also trigger the image significance determiner 40 determining that an additional representative image is required. Accordingly, Bozdagi discloses a parsing method and system for selecting a subset of representative of frames based on a comparison of adjacent frames, a comparison of an intensity of a frame against a threshold, or an expiration of a time period. However, these disclosures in Bozdagi have nothing to do with determining when there is a sufficient amount of caption data to begin associating the caption data with the audio-video signal. There is simply nothing in Bozdagi related to controlling the timing of the association of the caption data with the audio video signal. Additionally, Bozdagi is focused on image signals, not audio or caption data from the audio. If, as proposed by the Office, Alshawi is considered in
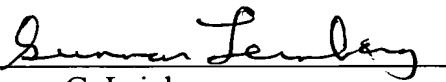
view of Bozdagi, at most the cited references would only disclose parsing the videophone conversation so that only certain frames of the videophone conversation were actually transmitted.

With the present invention, the rate at which the caption data is associated with the audio video signal is controlled to provide a quick and automated captioning of an audio and visual signal. One example of this process is illustrated in FIGS. 2, 4, and 5 and is described in paragraphs 28 and 32-37 in the above-identified patent application. Accordingly, in view of the foregoing amendments and remarks, the Office is respectfully requested to reconsider and withdraw the rejection of claims 3, 11, and 19.

In view of all of the foregoing, applicant submits that this case is in condition for allowance and such allowance is earnestly solicited.

Respectfully submitted,

Date: __May 10, 2004__

_____
Gunnar G. Leinberg
Registration No. 35,584

NIXON PEABODY LLP
Clinton Square, P.O. Box 31051
Rochester, New York 14603-1051
Telephone: (585) 263-1014
Facsimile: (585) 263-1600

CERTIFICATE OF MAILING OR TRANSMISSION [37 CFR 1.8(a)]

I hereby certify that this correspondence is being:

☒ deposited with the United States Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to: Mail Stop _____, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450

☐ transmitted by facsimile on the date shown below to the United States Patent and Trademark Office at (703) _____.

__May 10, 2004__
Date

_____
Signature

_____Sherri A. Moscato_____
Type or Print Name